

Monitoring the Simultaneous Presentation of Multiple Spatialized Speech Signals in the Free Field

W. Todd Nelson[†], Robert S. Bolia[✉], Mark A. Ericson[†], and Richard L. McKinley[†]

[†]*Air Force Research Laboratory, Wright-Patterson AFB, OH 45433 and* [✉]*Veridian, 5200 Springfield Street, Dayton, OH 45431*

Abstract: The effect of spatial auditory information on listeners' ability to detect, identify, and monitor multiple simultaneous speech signals was evaluated in the free field. Factorial combinations of three variables, including the number of localized speech signals, the location of the speech signals along the horizontal plane, and the sex of the talker were employed using a within-subjects design. Participants were required to detect the presentation of a critical speech signal among a background of non-signal speech events. Results indicated that the spatialization of simultaneous speech signals (1) increased the percentage of correctly identified critical signals and (2) lowered ratings of perceived mental workload as compared to a non-spatialized control condition.

INTRODUCTION

Advancements in the capabilities of modern tactical aircraft and weapons systems are likely to be associated with concomitant increases in the perceptual, perceptual-motor, and cognitive demands placed upon pilots. Accordingly, there is a compelling need to develop advanced interface concepts to compensate for these effects, thereby enabling crew members to operate effectively in these challenging environments. One potential way to offset the problems caused by these factors would be to exploit the human operator's ability to perceive and process spatial auditory information. Along this line, numerous researchers (1,2,3) have demonstrated that spatialized auditory displays increase performance efficiency for a variety of tasks that are relevant to airborne applications. In addition, displays that provide spatialized auditory information may also afford more efficient segregation, monitoring, and attentional shifts among auditory signals that are presented simultaneously. This notion is based, in part, upon the recognition that the spatial separation of acoustic signals improves the intelligibility of signals in noise and assists in the segregation of multiple sound streams - the so-called "cocktail party effect" (4). As noted by Wenzel (5), segregation enhancement may be particularly effective in applications involving both simultaneous speech channels - as is the case in aviation communications systems - and the kind of nonspeech auditory signals used in radar warning displays. Toward that end, the primary goal of this initial work was to determine the extent to which the spatialization of speech signals enhances the ability to detect, identify, and monitor multiple speech signals that are presented simultaneously in the free field.

METHOD

Participants. Four men and four women, naïve to the purposes of the experiment, served as paid participants. Their ages ranged from 20 to 50 years with a mean of 30 years. All participants had normal hearing and localization acuity.

Experimental Design. Five spatialization conditions (front right quadrant (RQ), front hemifield (FH), right hemifield (RH), full 360° (F), and a non-spatialized control (C)) were combined factorially with eight talker conditions (1,2,3,4,5,6,7, and 8 talkers) and the sex of the critical speech signal (male and female) to provide 80 experimental conditions. Participants completed all combinations of the experimental conditions with the constraint that each of the five spatialization conditions was performed during a separate experimental session.

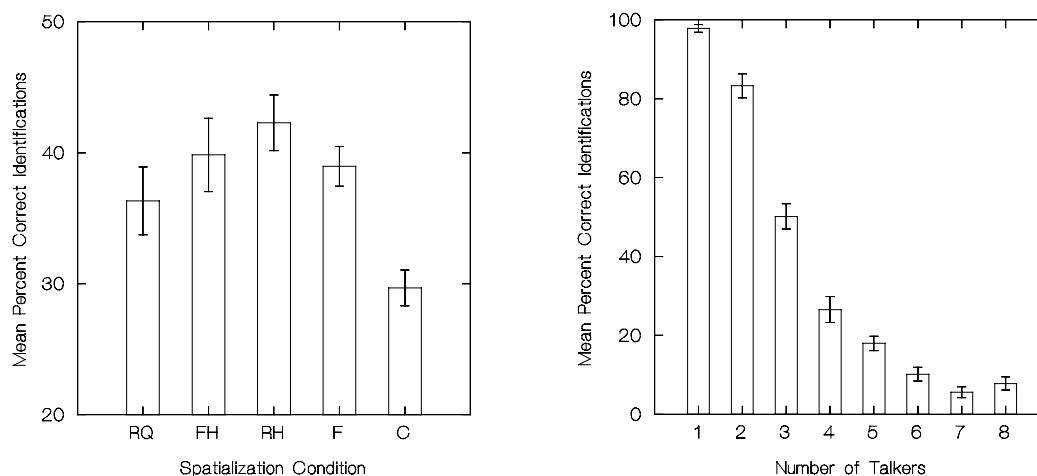
Apparatus. The experiment was conducted at the Air Force Research Laboratory's Auditory Localization Facility, Wright-Patterson Air Force Base, Ohio - a geodesic sphere that is housed within an anechoic chamber. In this investigation, auditory stimuli were restricted to the horizontal plane. Auditory speech signals consisted of a set of phrases from the coordinated call sign test (3) - a call sign (e.g. "Baron"), a color (e.g. "blue"), and a number (e.g. "two") embedded within a carrier phrase.

Procedure. Participants monitored the simultaneous presentation of multiple spatialized speech signals while seated in the center of the geodesic sphere. Their task was to listen for the occurrence of a critical call sign (e.g. "Baron") and to identify the color-number combination that emanated from the same spatial location as the critical call sign. For example, if assigned "Baron" as the critical call sign, then the appropriate response to "Ready Baron Go To Red Six Now" would have been to press the response buttons that corresponded to the red-six combination.

If the critical call sign was not presented, then no response was required. Fifty percent of the experimental trials included the critical call sign. Prior to data collection, participants completed five practice sessions, one practice session for each of the five spatialization conditions. Participants rated the perceived mental workload of the task by completing the NASA Task Load Index (6) after completing the non-spatialized control (C) and the full 360° (F) experimental sessions.

RESULTS

Percent Correct Identifications and Workload Ratings. Mean percentages of correct identifications were analyzed with a 5 (SPATIALIZATION conditions) x 8 (TALKER conditions) x 2 (SEX OF CRITICAL SIGNAL condition) repeated measures analysis of variance, which revealed that the SPATIALIZATION (see Figure 1) and TALKER (see Figure 2) main effects were statistically significant, $F(4,28) = 10.06$, $p < .05$ and $F(7,49) = 454.28$, $p < .05$, respectively. Post hoc pairwise comparisons of the SPATIALIZATION main effect indicated that performance efficiency associated with the four spatialization conditions was significantly greater ($p < .01$) than the non-spatialized control condition and that the four spatialization conditions were not significantly different from each other. An analysis of the workload data revealed that the FULL spatialization condition was associated with significantly lower ratings of workload than the non-spatialized control condition, $F(1,7) = 6.33$, $p < .05$. Collectively, these results provide a compelling demonstration of the beneficial effects of spatialization on performance efficiency and perceived mental workload and may be particularly pertinent to application domains in which operators are required to monitor effectively multiple speech communications channels simultaneously.



FIGURES 1 and 2. Figure 1 depicts the percentage of correct identifications under the five spatialization conditions. As can be seen in the figure, performance efficiency was superior in all the spatialization conditions (RQ, FH, RH, F) as compared to the non-spatialized control condition (C). Figure 2 illustrates mean percent correct identifications as a function of the number of simultaneous talkers. It is evident in the figure that performance efficiency varies inversely with the number of talkers. Error bars in Figures 1 and 2 denote the standard error of the mean.

ACKNOWLEDGMENTS

The authors wish to acknowledge the technical contributions of Michael L. Ward and Dennis L. Allen of Veridian.

REFERENCES

1. Bronkhorst, A.W., Veltman, J.A.H., & van Breda, L., *Human Factors*, **38**, 23-33 (1996).
2. Perrott, D.R., Cisneros, J., McKinley, R.L., & D'Angelo, W.R., *Human Factors*, **38**, 702-715 (1996).
3. McKinley, R.L., Ericson, M.A., D'Angelo, W.R., *Aviation, Space, and Environmental Medicine*, **5**, 31-38 (1994).
4. Bronkhorst, A.W., & Plomp, R., *Journal of the Acoustical Society of America*, **83**, 1508-1516 (1988).
5. Wenzel, E.M., *Presence*, **1**, 80-106 (1992).
6. Hart, S.G., & Staveland, L.E., In Hancock, P.A. & Meshkati, N. (Eds.), *Human Mental Workload*, North Holland, Elsevier Science Publishers, 1988, 139-183.